43-fold increase in anti-Muslim YouTube comments following Hamas' October 7 attack

By: Hannah Rose (https://www.isdglobal.org/isd_team/hannah-rose/) and Zahed Amanullah (https://www.isdglobal.org/isd_team/zahed-amanullah/)

19 December 2023

This article summarises the preliminary findings of analysis around anti-Muslim hate speech on YouTube, conducted using ISD and CASM Technology's bespoke anti-Muslim hate speech classifier, trained by analysts in the aftermath of the October 7 attack. We identified over 115,000 anti-Muslim comments from videos discussing the conflict between October 3 – 19 2023. This represents a 43-fold increase in the volume of anti-Muslim comments comparing the four days before and after October 7.

Anti-Muslim comments portrayed Islam as a source of hate, terrorism and evil; accused Muslims of propensity to lie; used dehumanising language against Muslims; and were associated with anti-Muslim Hindu ultra-nationalist ideologies.

Introduction

In the wake of the October 7 Hamas attack on Israel, communities globally have experienced increased isolation, polarisation and hate. A rise in anti-Muslim hate has been well evidenced; for example, UK-based Islamophobia monitoring charity Tell MAMA (https://tellmamauk.org/tell-mama-saw-more-than-a-seven-fold-increase-in-anti-muslim-cases-since-october-7/) recorded 1,350 cases of anti-Muslim incidents in the two months following October 7, a sevenfold increase on the same period the previous year. Over 700 of these cases occurred online, including on mainstream social media platforms and Telegram. Such anti-Muslim mobilisation is not limited to the UK, and attacks on Muslim or Palestinian individuals have been observed globally, including the killing of a 6-year-old boy (https://www.mirror.co.uk/news/us-news/man-paranoid-hamas-day-jihad-31205776?int_source=amp_continue_reading&int_medium=amp&int_campaign=continue_reading_button#amp-readmore-target) and the shooting of three Palestinian men (https://www.bbc.co.uk/news/world-us-canada-67539809) wearing kaffiyehs and speaking Arabic in the USA.

In this regard, previous ISD research (https://www.isdglobal.org/digital_dispatches/use-of-words-phrases-and-hashtags-associated-with-anti-muslim-mobilisation-surges-amid-israel-gaza-conflict/) presented an initial scoping of the online anti-Muslim hate landscape following October 7. Using keyword searches, analysts identified a 422% increase in keywords and phrases associated with anti-Muslim hate on X, and a tenfold increase on fringe platforms including Gab, Patriots.win and Bitchute. Anti-Muslim content promoted hate and spread negative stereotypes about Muslims, some of which was based on disinformation regarding the conflict.

To build on this previous analysis in a more nuanced fashion, the research presented in this briefing involved the construction of a bespoke hate speech classifier, in collaboration with CASM Technology, to facilitate the identification of hate speech at scale. This article outlines the initial findings of the classifier and content analysis of the over 115,000 anti-Muslim comments it identified in the immediate aftermath of October 7.

Methodology

In order to evaluate the volume of anti-Muslim hate in YouTube comments with a high degree of accuracy, ISD and CASM Technology constructed an automated hate speech classifier, similar to that built to measure online antisemitism (https://www.isdglobal.org/digital_dispatches/rise-in-antisemitism-on-both-mainstream-and-fringe-social-media-platforms-following-hamas-terrorist-attack/).

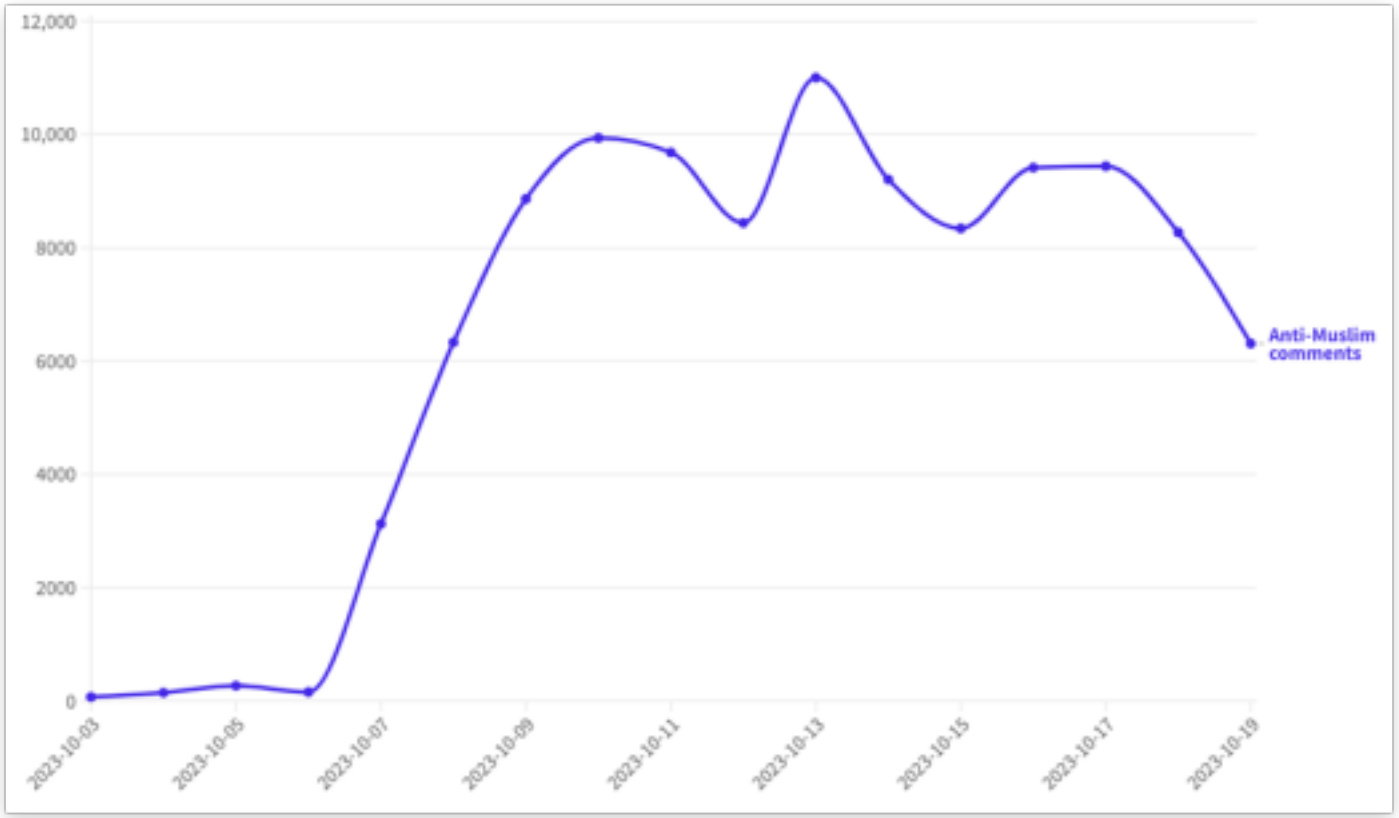
Anti-Muslim hate was defined using ISD's definition of targeted hate as '*activity which seeks to dehumanise, demonise, harass, threaten, or incite violence against an individual or community based on religion, ethnicity, race, sex, gender identity, sexual orientation, disability, national origin or migrant status*'. It was additionally recognised that anti-Muslim hate uses stereotypes or slurs about Muslims or Islam rooted in Islamophobia (https://www.isdglobal.org/explainers/islamophobia-anti-muslim-hatred/).

Two initial keyword lists were built and refined based on previous work on anti-Muslim hate; a first list of conflict-related keywords to identify relevant videos, and a second of those likely to be relevant to anti-Muslim hate in comments. These keyword lists identified over 13,000 videos about the October 7 attack and Israel-Gaza conflict, and 9.5 million comments on the videos were collected.

Groups of semantically similar messages were then clustered into groups of topics using natural language processing in order to achieve a qualitative overview of the relative prevalence of recurring themes within the dataset. Topics with over 500 relevant comments were assigned a broad qualitative theme using a random sample of 20 comments per topic, which are summarised in the second section of this article.

A full methodology is outlined in the methodological annex.

Data overview



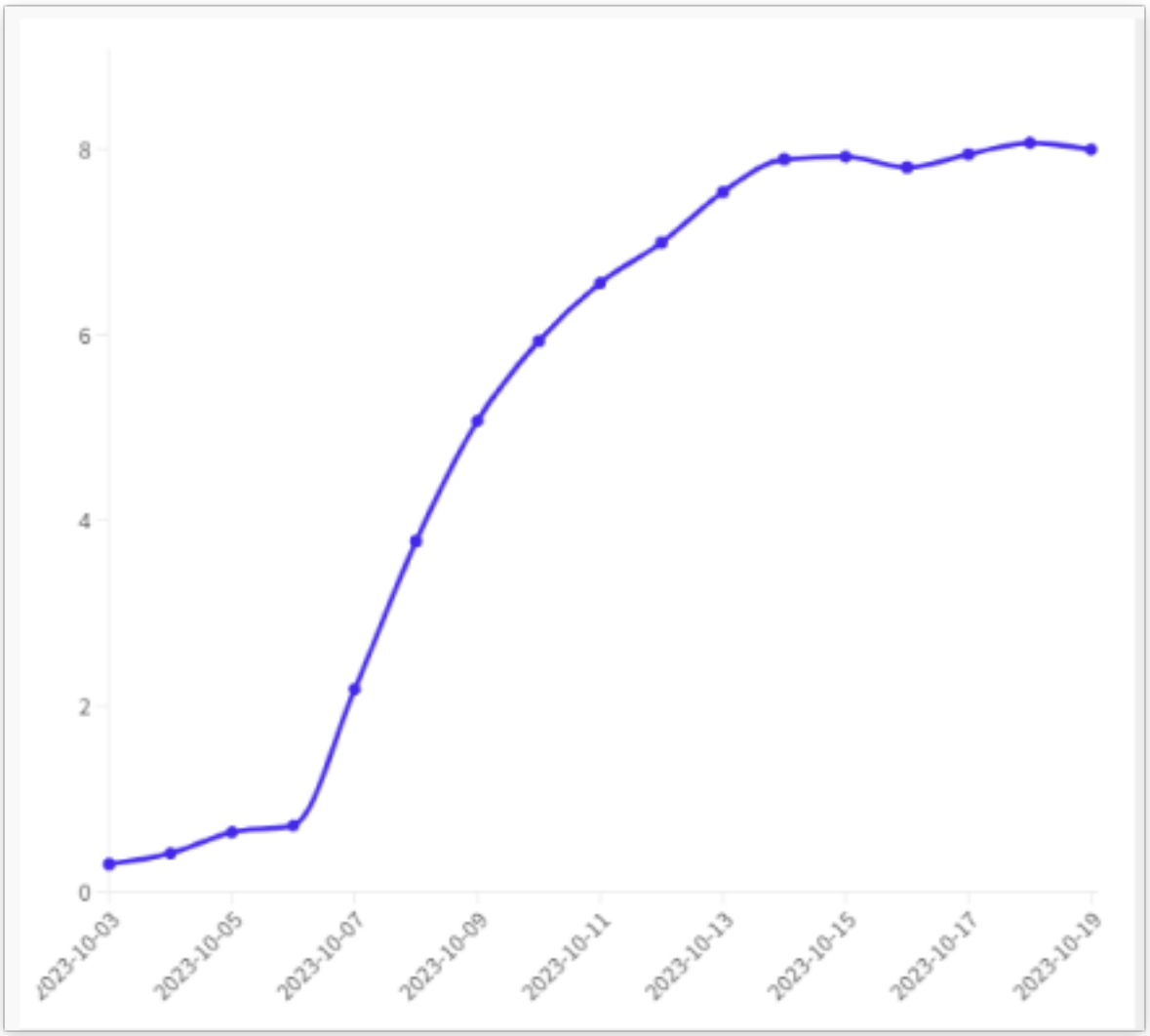
(https://www.isdglobal.org/digital_dispatches/43-fold-increase-in-anti-muslim-youtube-comments-following-hamas-october-7-attach/graph-1-2/)

Graph 1: The volume of anti-Muslim YouTube comments over time.

On videos about the conflict between 3 and 19 October, the classifier identified over 115,000 anti-Muslim YouTube comments. Comparing the four days before and after October 7, there was a 43-fold increase in the volume of anti-Muslim comments. Anti-Muslim hate comments rose sharply in the aftermath of October 7, and remained at an elevated level for the following 12 days.

While few conclusions can be drawn on user demographics, the presence of such a volume of anti-Muslim hate on a mainstream platform demonstrates the wide availability of harmful content, which is not confined to fringe ecosystems or the political extremes. Anti-Muslim attitudes are apparent across the political spectrum, including among those who may not otherwise hold extremist attitudes.

The highest number of anti-Muslim comments on one day (11,004) was collected on October 13, coinciding with the so-called 'Day of Rage' motivated by former Hamas leader Khaled Meshal (https://www.reuters.com/world/middle-east/former-hamas-chief-calls-protests-neighbours-join-war-against-israel-2023-10-11/) in response to Israeli retaliatory bombing of the Gaza strip. A video of Meshal's call for demonstrations across 'the Islamic world' was widely shared online (https://www.isdglobal.org/digital_dispatches/use-of-words-phrases-and-hashtags-associated-with-anti-muslim-mobilisation-surges-amid-israel-gaza-conflict/) and interpreted as a heightened threat among Western communities on October 13. While protests did occur (https://abcnews.go.com/International/hamas-day-rage-protests-break-middle-east/story?id=103955873) and communities in Western countries expressed justifiable concerns, the call was also manipulated (https://www.isdglobal.org/digital_dispatches/use-of-words-phrases-and-hashtags-associated-with-anti-muslim-mobilisation-surges-amid-israel-gaza-conflict/) by anti-Muslim actors as a 'Day of Jihad', spreading the conspiracy that Muslims across the world would rise up and turn on their neighbours. This spread of anti-Muslim messaging relating to the so-called 'Day of Rage' has presented a clear impact in the volume of anti-Muslim hate on October 13.



(https://www.isdglobal.org/digital_dispatches/43-fold-increase-in-anti-muslim-youtube-comments-following-hamas-october-7-attach/graph-2-2/)

Graph 2: The proportion of cumulative anti-Muslim comments per cumulative number of videos collected.

Given that a rise in the absolute volume of anti-Muslim comments may be impacted by an increase in the number of videos relevant to the conflict since October 7, a second graph shows the proportion of cumulative anti-Muslim comments per cumulative number of videos collected by the classifier. Once again, this graph depicts a steep rise following October 7. Overall, the data show an eightfold increase in the proportion of cumulative anti-Muslim comments on cumulative videos about Israel and Palestine comparing before and after October 7. This demonstrates that although the overall number of videos and comments collected after October 7 increased, so did the proportion of anti-Muslim hate comments per video.

Semantic analysis

From the dataset of anti-Muslim comments, groups of semantically similar messages were clustered in order to achieve a qualitative overview of the relative prevalence of different recurring themes in the dataset, resulting in a list of 98 distinct topics.

In order to produce a preliminary analysis, each message cluster that comprised of 500 posts or more – 26 of 98 – was analysed by subject matter experts to categorise them according to a number of themes common to widely used anti-Muslim narratives. These 26 message clusters consisted of over 48,000 comments, or nearly 80% of total comments allocated to topics.

From these topics came various distinct anti-Muslim themes. The most common themes, as might be expected, portrayed Islam as providing justification for terrorism, hate, and evil toward Israel and the world, with Muslims willing perpetrators. These themes alone constitute nearly half the comments of all analysed clusters. Other notable themes include accusing Muslims of a propensity to lie (*taqqiya*), to engage in genocide, and to specifically target Jews with hatred. Lesser used themes engaged in dehumanising language towards Muslims (using words like "animals"), insisted Muslims could only subjugate others, and were violent and misogynistic by nature.

Some specific comment examples for the most common themes include the following:

- Islam as an enabler of terrorism: "Muslims, Islam and terrorism are one and the same" and "Its ok. Allaha has been buried here, under the demolished mosque (the breeding point of terrorists)."
- Islam as a blueprint for global conquest: "This was designed and executed with precision to destroy the West and implement global fascism" and "Meanwhile every Globalist government in the West is inviting millions of these cockroaches into our ancient homelands and they will never stop."
- Islam as inherently antisemitic: "Zero Muslims condemn Hamas because they quietly and secretly agree and want all Jews gone forever all over the world."
- Genocidal language against Muslims: "Israel is totally justified if they decide to Carpet bomb every square inch of Palestine to eliminate the terrorists of Islam" and "As much as I hate the CCP their method for eradicating Muslim extremism has worked the best so far."
- Muslims are taught to lie: "So, all this pearl clutching and taqqiya by Muslims is hard to take seriously" and "I've read enough of the Koran to know it teaches to devalue the worth of the Kefir, and that every word out of any Muslims mouth is Taqqiyah."

Beyond these themes, consisted of topics explicitly associated with Hindutva (https://www.isdglobal.org/explainers/hindutva-hindu-nationalism/) narratives, a form of Hindu ultra-nationalism which often adopts anti-Muslim beliefs. Here, promoters of anti-Muslim narratives from a separate regional conflict between Hindutva movements and Indian Muslims attempted to draw parallels and connections with the Israel-Hamas conflict. While categorised by regional affiliation in this instance, these comments also referenced many of the other themes tracked here, including framing Islam as inherently violent or a conquering force.

This suggests an amplification potential not directly related to the Israel-Hamas conflict or, more broadly, a potential for exploiting anti-Muslim discourse in one region to endorse similar conflicts in another. There is also the possibility of amplification by state or other malign actors for other purposes, as has been the case with the Israel-Gaza war more broadly (https://www.isdglobal.org/digital_dispatches/capitalising-on-crisis-russia-china-and-iran-use-x-to-exploit-israel-hamas-information-chaos/), though that was not the focus of this study.

Conclusion

This research has quantitively evidenced a rise in online anti-Muslim hate in the aftermath of the October 7 Hamas attack, and a sustained increase throughout the subsequent Israel-Gaza war. Geopolitical events and crises such as the current war have demonstrable impacts on faith and minority communities across the globe, including a rise in hatred, polarisation and fear.

YouTube's Hate Speech policy (https://support.google.com/youtube/answer/2801939?hl=en) prohibits

content which may 'incite hatred against individuals or groups based on their protected group status', and that such content will be removed. Specifically, it gives examples of prohibited hate speech, including '[People with protected group status] are dogs' or '[people with protected group status] are like animals.'

The data gathered for this report demonstrates the volume of anti-Muslim hate, much of which is likely to directly contravene Terms of Service, which is slipping through the net of moderation. While anti-Muslim hatred can take covert or coded forms, it is – as ISD research (https://www.isdglobal.org/digital_dispatches/unpacking-antisemitic-narratives-about-the-hamas-attack-and-the-israel-hamas-conflict-within-youtube-comments/) previously showed of antisemitism – often overtly hateful, in clear breach of platform policies and could be illegal. Platforms across the mainstream and fringe ecosystems have failed (https://www.isdglobal.org/digital_dispatches/use-of-words-phrases-and-hashtags-associated-with-anti-muslim-mobilisation-surges-amid-israel-gaza-conflict/) to effectively moderate anti-Muslim hate speech, and must do more to implement their own Terms of Service. This will become ever more urgent in the context of incoming digital regulation regimes in the UK and EU, where platforms will have legal duties to remove content which is explicitly illegal.

Anti-Muslim hate has long suffered from inadequate data collection, where few governments or organisations systematically collect incident reports, and very little understanding exists of its online proliferation. More monitoring, both of online hate speech and offline incidents, is vital to build a full picture of the anti-Muslim landscape, particularly in the current context of polarisation and rising hate towards Muslims.

*****

METHODOLOGICAL ANNEX

Video collection

A first keyword list was created based on ethnographic research to identify videos which would be relevant to October 7 and the subsequent Israel/Gaza war, and 13,639 videos were collected from 3 to 19 October. From these videos, 9,560,938 comments were collected.

Filtering

Given that the vast majority of comments were not anti-Muslim, filtering was a necessary next step in order to improve the practicality of labelling and evaluating training data, and the functioning of Natural Language Processing (NLP) models. Firstly, all comments were put through an English language annotator to identify English messages – also removing messages that are nonsensical or just emojis – leaving 5,746,192 messages.

Keyphrase lists were then created using both ethnography and previous lists of anti-Muslim slurs and keywords and phrases, leading to 12 keyword lists in total, including 3 generic hateful annotations. A round of topic modelling was carried out to identify topic with potential relevance.

Where analysts were uncertain on the relevance of certain keywords, manual methods were employed to determine their relevance. For example, a manual review of mentions of 'jihad' in 200 comments found 6 anti-Muslim mentions and 8 non-anti-Muslim mentions, leading analysts to retain its inclusion as a keyword. Additionally, after a large volume of messages were identified regarding Islam Makhachev – due to the aggressive language which often arises in discussion of boxing – an effort was made to remove comments relating to sport to filter out this noise. Additionally, analysts reviewed a random sample of 200 comments which had been filtered out by the collection in order to determine that words or phrases of relevance were not being omitted. After filtering, 269,718 comments were identified as potentially relevant to anti-Muslim

hate.

Labelling training data

Next, analysts manually labelled 2,000 pieces of training data against their relevance to both anti-Muslim and anti-Islam hatred, with 0 being not relevant, 1 being edge cases, and 2 as a high likelihood of relevance. A random sample of each analyst's data subset was blind second coded by a peer, returning a high level of inter-coder reliability.

Therefore, anti-Islam comments which did not contain anti-Muslim hatred were only labelled as relevant where they overtly adopted anti-Muslim framings, and all edge cases which were labelled as 0 anti-Muslim and 1 anti-Islam were resolved.

Ensemble construction

The ensemble comprised of 60 annotations and 28 NLP models, including 3 trained annotators on annotated data created by ISD for this project. Of the 2,000 pieces of training data, 799 comments were used to train the annotators, 599 of which were used for training and 200 for evaluation. The remaining 1,201 labelled comments were used to train and evaluate the model – 961 for training and 240 for evaluation.

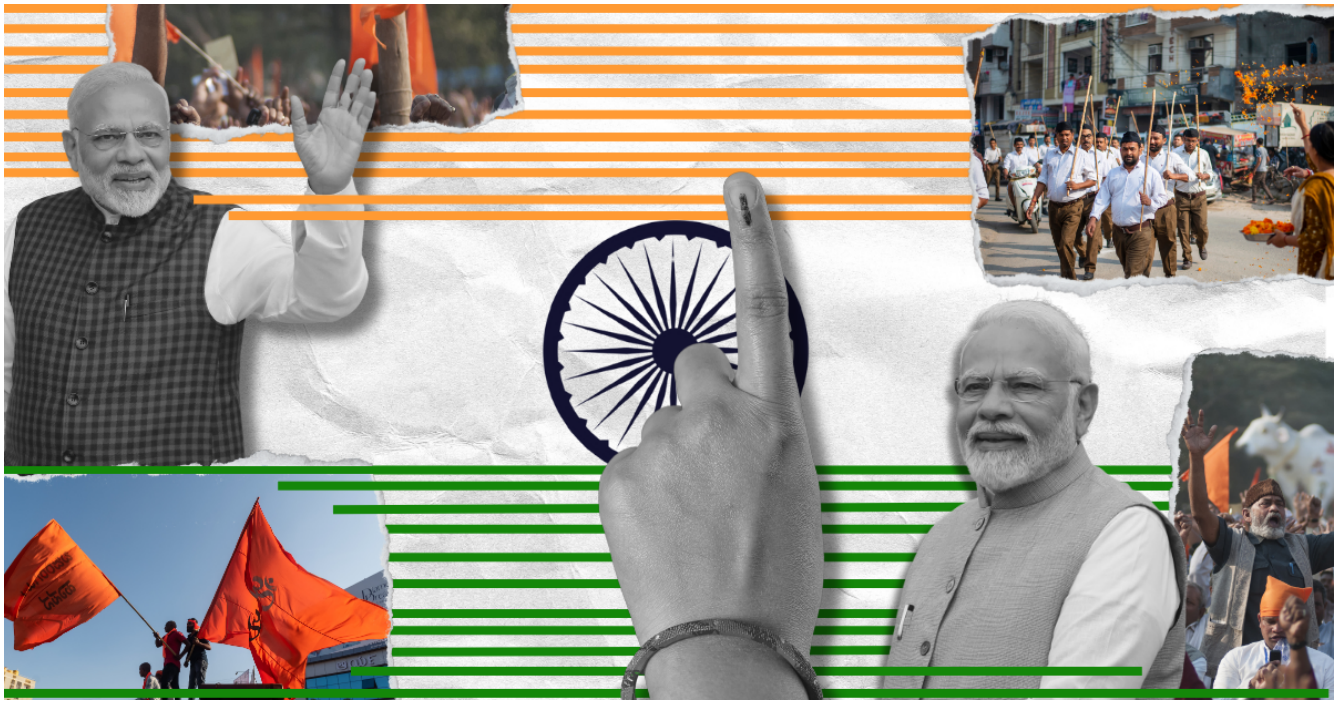
This resulted in a dataset of 115,985 anti-Muslim comments. The final model had precision, recall and F1 weighted averages of 0.82.

Topic modelling

Topic modelling is a powerful method for identifying recurring discussions, sentiments and themes in large-scale text data, and patterns in how these discourses are used. Comments were excluded from topic modelling where they were too long. Topic Modelling was used to analyse the dataset of 113,666 probable anti-Muslim comments from selected YouTube videos, using the machine learning and natural language processing tool, known as BERTopic. This enabled the automated mapping and identification of 'topics', which are distinct clusters of messages that occur within a dataset based on their linguistic similarity. Topics with over 500 related comments were labelled by their relevance based on a random sample of 20 related comments. Topics were then qualitatively grouped into themes.

With thanks to Shaun Ring (CASM) for classifier construction, Kieran Young (CASM) for data visualisation and Jon Jones (CASM) for topic modelling; and ISD colleagues Zoe Manzi, Guy Fiennes, Michel Seibriger and Jacob Davey for labelling training data.

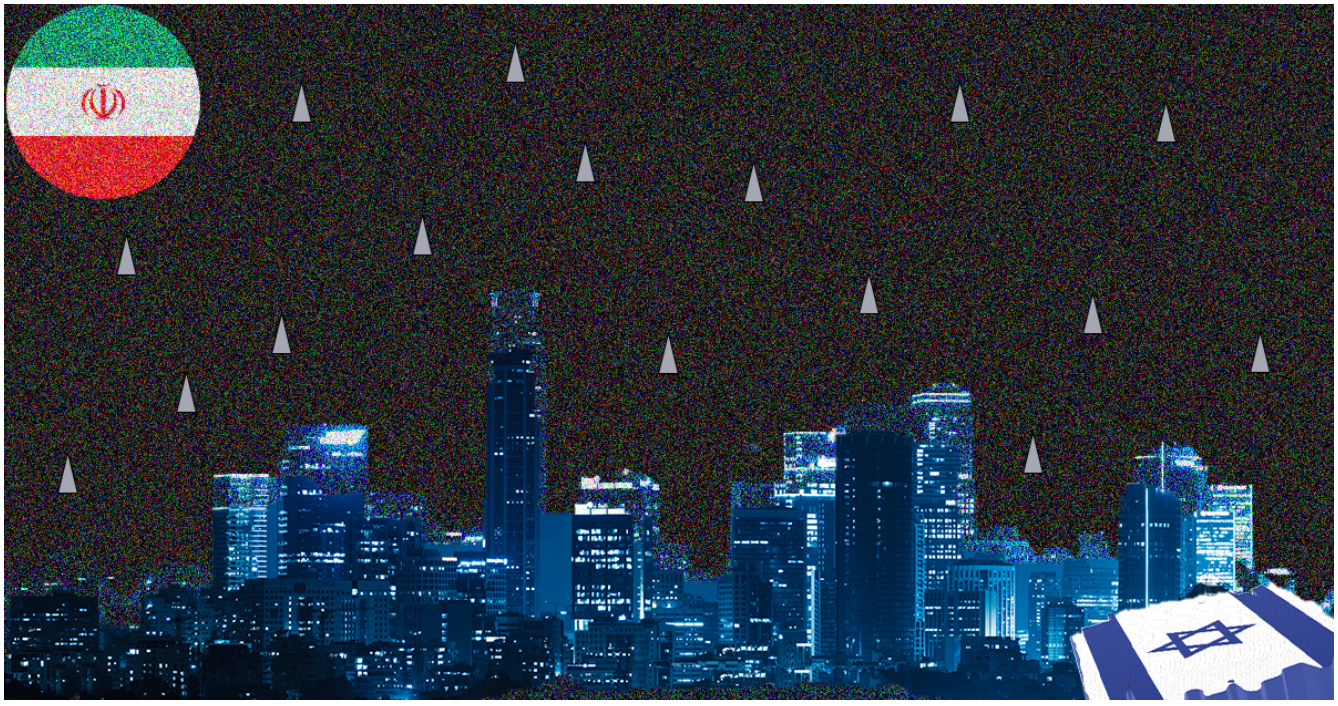
Related



(https://www.isdglobal.org/digital_dispatches/hindu-nationalists-push-conspiracy-theories-and-misinformation-ahead-of-elections/)

Hindu nationalists push conspiracy theories and misinformation ahead of elections (https://www.isdglobal.org/digital_dispatches/hindu-nationalists-push-conspiracy-theories-and-misinformation-ahead-of-elections/)

India's upcoming general elections are likely to hand the Hindu nationalist Bharata Janata Party (BJP) a third term in office. This Dispatch examines Hindutva's use of divisive rhetoric and conspiracy theories ahead of the election.



(https://www.isdglobal.org/digital_dispatches/misleading-and-manipulated-content-goes-viral-on-x-twitter-in-middle-east-conflict-iran-israel-strikes/)

Misleading and manipulated content goes viral on X in Middle East conflict (https://www.isdglobal.org/digital_dispatches/misleading-and-manipulated-content-goes-viral-on-x-twitter-in-middle-east-conflict-iran-israel-strikes/)

Within seven hours of Iranian drones being launched towards Israel, misleading content received over 37 million views on X.

(https://www.isdglobal.org/digital_dispatches/pro-ccp-spamouflage-campaign-experiments-with-new-tactics-targeting-the-us/)

Pro-CCP Spamouflage campaign experiments with new tactics targeting the US (https://www.isdglobal.org/digital_dispatches/pro-ccp-spamouflage-campaign-experiments-with-new-tactics-targeting-the-us/)

The pro-CCP network known as Spamouflage is a long-running and widespread but largely ineffective operation. However, there is evidence that the network is now experimenting with new ways to increase their influence.

(https://www.facebook.com/ISDglobal) (https://twitter.com/ISDglobal) (https://www.linkedin.com/company/isdglobal/)

ISD

Amman | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2024). Institute for Strategic Dialogue (ISD) is a company limited by guarantee, registered office address PO Box 75769, London, SW1P 9ER. ISD is registered in England with company registration number 06581421 and registered charity number 1141069. All Rights Reserved. Any copying, reproduction or exploitation of the whole or any part of this website without prior written approval from ISD is prohibited.